

ANALISIS KLASTERISASI JUMLAH PENDERITA PENYAKIT MENGGUNAKAN K-MEANS SEBAGAI DASAR DISTRIBUSI LAYANAN RUMAH SAKIT UMUM DI SUMATERA SELATAN

Siti Hariza Marshella¹, Naila Raihana Putri², Eka Therina Lakeisyah^{*3}, M. Fadhil Rahman⁴, Aliyananda Risyahputri⁵, Rahmat Maulana⁶, Ken Ditha Tania⁷, Winda Kurnia Sari⁸

¹⁻⁸Program Studi Sistem Informasi, Universitas Sriwijaya, Palembang

email: ¹sitiharizamarshella@gmail.com, ²nailaraihana55511@gmail.com, ³ekathrnl@gmail.com,

⁴fadhilr637@gmail.com, ⁵aliyanandar04@gmail.com, ⁶maulanaaa.rahmat@gmail.com, ⁷kenya.tania@gmail.com, ⁸windakurniasari@unsri.ac.id

Abstract

The increasing number of disease cases in recent years in South Sumatera Province has impacted the evaluation of equitable healthcare service distribution across all regions. This study highlights the disparity in healthcare facilities, particularly public hospital services, in regencies and cities in South Sumatera Province. The objective of this research is to furnish the government with data-based insights to serve as a reference point in the formulation of policies that distribute health services in a more equitable. This study utilizes a dataset from Badan Pusat Statistik (BPS) South Sumatera, processed using the K-Means algorithm through RapidMiner and Python. The data analysis results classify regencies/cities into three clusters: Cluster 0 (low) consists of 11 regions, Cluster 1 (high) consists of 1 region, and Cluster 2 (moderate) consists of 5 regions. The clustering results indicate an imbalance in healthcare facility distribution, particularly between healthcare services in Palembang City and other regencies/cities. Based on these findings, it is recommended that the government revise the policies regarding the equitable distribution of healthcare services and medical personnel in each region. Additionally, the adoption of healthcare service innovations through a knowledge management approach could optimize the equalization of healthcare services in South Sumatera.

Keywords: Knowledge Discovery, K-Means Clustering, Healthcare Distribution

Abstract

Peningkatan jumlah penderita penyakit dalam beberapa tahun terakhir pada Provinsi Sumatera Selatan berimbas kepada evaluasi pendistribusian layanan kesehatan yang merata di setiap daerah. Penelitian ini menyoroti ketidakmerataan fasilitas kesehatan berupa layanan rumah sakit umum yang tersebar di masing-masing wilayah yang ada di Provinsi Sumatera Selatan. Tujuan dari penelitian ini adalah untuk memberikan wawasan berbasis data sebagai acuan bagi pemerintah dalam pengambilan kebijakan pendistribusian layanan kesehatan agar dapat lebih merata. Penelitian ini menggunakan dataset yang berasal dari BPS Provinsi Sumatera selatan dan diolah dengan algoritma K-Means melalui rapid miner dan python. Hasil dari analisis data tersebut adalah mengelompokkan wilayah Kabupaten/Kota kedalam 3 kluster yakni kluster 0 (rendah) terdiri dari 11 wilayah, kluster 1 (tinggi) terdiri dari 1 wilayah, dan kluster 2 (sedang) terdiri dari 5 wilayah. Interpretasi dari klasterisasi dan pengolahan data menunjukkan adanya ketimpangan dalam pendistribusian fasilitas kesehatan terutama antara layanan kesehatan di wilayah Kota Palembang dengan Kabupaten/Kota lainnya. Sehingga, dari temuan tersebut direkomendasikan bagi pemerintah untuk melakukan kebijakan ulang terkait pendistribusian layanan dan tenaga kesehatan di tiap daerah secara merata dan dapat menerapkan inovasi layanan kesehatan dengan pendekatan knowledge management yang dapat mengoptimalisasi pemerataan layanan kesehatan di Sumatera Selatan.

Kata Kunci: Knowledge Discovery, K-Means Clustering, Distribusi Layanan Kesehatan

1. PENDAHULUAN

Kesehatan masyarakat berperan selaku salah satu standar tingkat kemajuan pembangunan manusia (Bustamam *et al.*, 2021). Menurut *United Nations Development Programme* (UNDP) indeks pembangunan manusia merupakan salah satu alat untuk mengukur angka kesejahteraan suatu daerah atau negara (Yulianti &

<https://doi.org/10.35145/joisie.v9i1.4959>

JOISIE licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

Qomariah, 2025). Sehingga, berbagai upaya terus dilakukan pemerintah untuk meningkatkan kualitas layanan kesehatan terhadap masyarakat, termasuk pemerintah wilayah Sumatera Selatan.

Peningkatan jumlah kasus penderita penyakit di Sumatera Selatan dalam beberapa tahun terakhir menurut Dinas Kesehatan Provinsi Sumatera Selatan pada dinkes.sumselprov.go.id menjadi perhatian serius, terutama upaya pemerintah dalam menyediakan layanan kesehatan yang adil dan merata. Berdasarkan data Dinas Kesehatan Provinsi Sumatera Selatan (2023), terjadi lonjakan kasus penyakit dari 630.377 kasus pada tahun 2019 menjadi 3.261.949 kasus pada tahun 2023. Persentase kenaikan kasus per tahun-nya pun menunjukkan tren yang mengkhawatirkan. Terjadi peningkatan sebesar 50,69% kasus dari tahun 2019 ke tahun 2020. Lalu, pada tahun 2020 ke tahun 2021 terjadi peningkatan sebesar 52,82% kasus. Pada tahun 2021 ke tahun 2022 terjadi pembengkakan kasus menjadi 79,45% kasus, dan pada tahun 2022 ke tahun 2023 terjadi peningkatan sebesar 25,22% kasus.

Meningkatnya kurva jumlah kasus penderita penyakit tersebut menuntut pemerintah untuk menyediakan fasilitas pelayanan yang setara. Ironisnya, dalam data yang diambil dari Badan Pusat Statistik Provinsi Sumatera Selatan tahun 2023 dalam sumsel.bps.go.id, terlihat jelas bahwasanya terdapat ketimpangan yang nyata dalam alokasi fasilitas kesehatan di berbagai kabupaten/kota. Kabupaten Ogan Ilir mencatat jumlah penderita penyakit sebanyak 16.321 kasus, namun hanya memiliki tiga rumah sakit umum dan 1.818 tenaga kesehatan yang mana jumlah ini lebih rendah dibandingkan daerah dengan jumlah penderita lebih sedikit. Selain itu, Kota Palembang sebagai ibukota provinsi memiliki 23 rumah sakit umum dan 10.411 tenaga kesehatan, yang mengindikasikan bahwa fokus pemerintah pada layanan kesehatan masih terpusat pada kota besar (Badan Pusat Statistik Provinsi Sumatera Selatan, 2023b). Kondisi ini berpotensi memperburuk akses masyarakat di daerah dengan layanan yang terbatas dalam meraih kesempatan yang sama untuk memenuhi hak kesehatannya sebagai warga negara Indonesia.

Menindaklanjuti urgensi tersebut, pengelompokan wilayah berdasarkan jumlah kasus penderita penyakit dengan menerapkan teknik clustering pada data mining menggunakan algoritma *machine learning K-Means* hadir sebagai salah satu solusi. *Data mining* merupakan sebuah upaya untuk mengidentifikasi nilai tambah pada sebuah data yang belum terungkap menggunakan statistika matematika (Setyaningtyas et al., 2022). Pada proses *clustering*, *data mining* berperan dalam *men-tracking* motif pada sebuah kumpulan data yang tersimpan dalam jumlah besar (Nirwana Sevi Dian, 2022). Data tersebut lalu akan dikelompokkan ke dalam beberapa *cluster* berdasarkan pola kesamaan jumlah penderita penyakit, ketersediaan rumah sakit umum, dan jumlah tenaga kesehatan di setiap daerah. Hasil akhir dari clustering ini akan mengidentifikasi wilayah prioritas di Sumatera Selatan yang membutuhkan perbaikan layanan rumah sakit umum serta dipetakan ke dalam saran kebijakan pemerintah sebagai bentuk penerapan *knowledge management*.

Pemilihan algoritma K-Means dalam penelitian ini didasarkan pada efisiensinya dalam menangani dataset berukuran besar dan kemampuannya untuk menghasilkan segmentasi yang jelas berdasarkan variabel numerik. Meskipun terdapat algoritma clustering lain seperti DBSCAN atau hierarchial clustering, K-Means tetap menjadi pilihan utama dalam banyak studi karena kesederhanaannya dan kecepatan komputasi yang tinggi (Hendriansyah et al., 2025; Mayola et al., 2025). Selain itu, penelitian oleh Gesicho et al.(2021) menunjukkan bahwa K-Means efektif dalam mengevaluasi kinerja fasilitas kesehatan dalam memenuhi persyaratan pekaporan indikator HIV di Kenya, memberikan dasar yang kuat untuk pengambilan keputusan dalam perencanaan layanan kesehatan. Hal serupa juga ditunjukkan oleh Sitinjak et al.(2022) yang berhasil mengelompokkan kecamatan di kabupaten Karawang berdasarkan jumlah tenaga kesehatan. Hasilnya mengungkapkan ketimpangan distribusi tenaga medis antar wilayah, yang kemudian digunakan sebagai acuan perencanaan pemerataan tenaga kesehatan di daerah tersebut.

Metode *clustering* telah banyak digunakan untuk menganalisis pola penyakit dan mendukung perencanaan layanan kesehatan. Ordila et al. (2020) menerapkan *K-Means* untuk mengelompokkan data catatan kesehatan pasien menurut jenis penyakit dengan tingkat akurasi *clustering* sebesar 85%. Hasil penelitian ini membantu klinik dalam memahami distribusi penyakit dan meningkatkan efisiensi layanan kesehatan dengan menyediakan data yang lebih terstruktur untuk pengambilan keputusan. Selain itu, penelitian oleh Minarni et al. (2021) menggunakan *K-Medoids* dalam pengelompokan penyakit dan menemukan bahwa metode ini memiliki stabilitas klaster yang lebih tinggi dibandingkan *K-Means* dalam beberapa kasus, dengan peningkatan akurasi clustering sebesar 5%. Output dari penelitian ini adalah rekomendasi distribusi sumber daya medis berdasarkan pola penyakit yang teridentifikasi. Di lain sisi, Vira (2021) membandingkan beberapa teknik clustering, seperti *K-Means*, *K-Medoids*, dan *Fuzzy Clustering*, dan menemukan bahwa *Fuzzy Clustering* memiliki performa terbaik dalam menangani data dengan pola yang tidak teratur, dengan akurasi *clustering* mencapai 87%. Hasil penelitian ini menghasilkan pemetaan risiko penyakit di berbagai provinsi di Indonesia, yang dapat digunakan untuk meningkatkan efektivitas intervensi kesehatan.

<https://doi.org/10.35145/joisie.v9i1.4959>

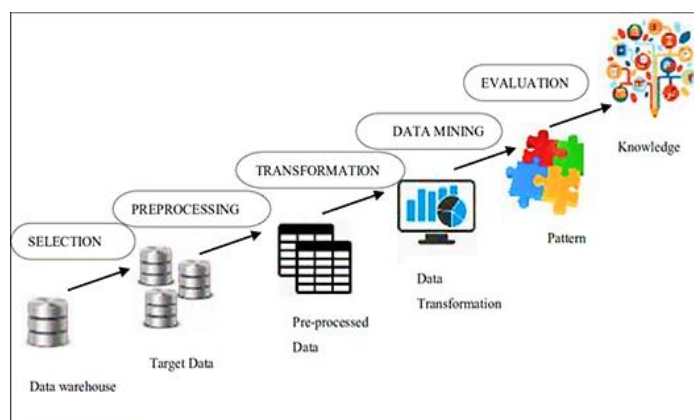
JOISIE licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

Lain halnya dengan penelitian terdahulu, fokus penelitian ini akan bertumpu pada empat variabel yakni jumlah penduduk, jumlah rumah sakit, jumlah penderita penyakit, serta jumlah tenaga kesehatan yang bersumber dari 17 kabupaten/kota di Sumatera Selatan. Poin lainnya yang juga memberikan diferensiasi adalah output akhir dari *clustering* dengan algoritma *K-Means* yang berupa solusi kebijakan bagi pemerintah provinsi Sumatera Selatan. Dengan demikian, diharapkan penelitian ini tidak hanya memberikan gambaran pengelompokan mengenai jumlah kasus penderita di tiap daerah, namun juga mendukung pengambilan kebijakan pemerintah berbasis data yang relevan demi distribusi pelayanan rumah sakit umum yang berkeadilan bagi masyarakat di luar pusat kota. Bagian selanjutnya akan membahas mengenai metodologi serta tahapan di dalam penelitian, analisis hasil serta pembahasan dari proses *clustering*, analisis pengusulan kebijakan, serta kesimpulan dan saran yang dapat dapat diterapkan pada inovasi penelitian futuristik lainnya.

2. METODE PENELITIAN

2.1 TEKNIK PENGUMPULAN DATA DAN SUMBER DATA

Fase pertama, data-data sekunder yang bersumber dari dokumen atau laporan resmi yang tersedia dihimpun dengan mengimplementasikan teknik dokumentasi. Pendekatan Knowledge Discovery in Database (KDD) diterapkan dalam proses *clustering* terhadap data tersebut. Pendekatan ini mencakup serangkaian tahapan yang memanfaatkan hasil penelusuran pola pada data yang dikumpulkan, sehingga menjadi sebuah informasi yang dapat dicerna dengan mudah (Utomo et al., 2023). Nantinya informasi yang diperoleh dari proses KDD dapat menjadi sebuah database yang berperan selaku landasan dalam mengambil keputusan (Atalya et al., 2024).



Gambar 1. Tahapan Knowledge Discovery in Database

Merujuk pada penelitian oleh Leza *et al.* (2024) tahapan dalam KDD dibagi menjadi lima tahapan utama yakni penyeleksian data, *data pre-processing*, *data transformation*, penambangan data dan evaluasi hasil seperti yang digambarkan di Gambar 1. Langkah pertama sebelum data diolah menjadi sebuah informasi yaitu dengan memilih data-data tertentu dari sebuah basis data besar, lalu disimpan secara terpisah. Setelah itu, data akan melalui tahapan *pre-processing* atau pembersihan serta penambahan informasi eksternal apabila diperlukan. Dengan kata lain, keseluruhan data yang teridentifikasi duplikat, memiliki nilai yang inkonsisten atau ekstrim akan dihapus untuk memaksimalkan hasil yang dituju. Selanjutnya, data-data terpilih tersebut akan ditransformasikan agar selaras dengan proses penambangan data yang ingin dilakukan. Umumnya, langkah *transformation* ini berupa normalisasi serta pemilihan fitur atau *feature selection* yang mewakili data serta tujuan penelitian. Masuk ke dalam proses utama yakni *data mining*, pola dari data yang dipilih akan dilacak kecenderungannya. Pada proses inilah, informasi yang terpendam akan digali sedemikian rupa menggunakan algoritma tertentu tergantung penelitian, yang mana dalam penelitian ini menggunakan *machine learning* teknik *clustering*. Apabila proses data mining telah dilakukan, maka proses KDD masuk pada tahapan akhirnya yakni *evaluation*. *Evaluation* merupakan tahap menginterpretasikan kesesuaian pola yang didapatkan dengan fakta atau paham yang sudah ada. Tahap ini juga menghasilkan pengetahuan yang dapat diterapkan dalam proses *decision making*.

2.2 DATA SELECTION (PENGUMPULAN DATASET)

Tabel 1. Atribut Data

No	Basis Data	Atribut
1	Data BPS	Kabupaten/Kota
2		Jumlah Penduduk
3		Jumlah Rumah Sakit Umum
4		Jumlah Penderita Penyakit
5		Jumlah Tenaga Kesehatan

Merujuk pada Tabel 1, dataset yang dimanfaatkan dalam penelitian ini diperoleh dari sumber data instansi resmi, yakni Badan Pusat Statistik Provinsi Sumatera Selatan (2023). Dataset ini merupakan data time-series (2019–2023) yang dikumpulkan per kabupaten/kota, sehingga mencerminkan perkembangan setiap atribut dari waktu ke waktu di masing-masing kabupaten/kota. Data yang dikumpulkan (atribut data) meliputi :

- Jumlah penduduk di setiap kabupaten/kota di Provinsi Sumatera Selatan.
- Jumlah rumah sakit umum yang tersebar di wilayah tersebut.
- Jumlah penderita penyakit di setiap kabupaten/kota.
- Jumlah tenaga kesehatan di setiap daerah.

Tujuan dari pengumpulan dataset ini adalah untuk memberikan gambaran menyeluruh terkait pola distribusi penyakit yang kedepannya akan digunakan dalam alur pengelompokan menggunakan model K-Means Clustering sebagai acuan dan saran bagi pemerintah untuk distribusi rumah sakit umum di tiap daerah di Provinsi Sumatera Selatan.

2.3 DATA PRE-PROCESSING

Setelah dataset dikumpulkan, tahapan selanjutnya adalah tahap *pre-processing*, pada tahap ini dataset dianalisis untuk mengekstraksi atribut yang relevan dan menentukan *data training* dan *data testing* dari dataset. Tahapan *pre-processing* data dinilai krusial untuk memastikan kualitas data, meningkatkan performa model dalam rangka menghasilkan hasil yang akurat (V. Novalia et al., 2024). Untuk sampel, peneliti mengambil fitur yang relevan untuk dianalisis. Pengujian dapat dimulai dengan melakukan *split* dataset menjadi data latih dan uji dengan proporsi 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, dan 10:90 (Mohammed & Alsunosi, 2022). Dalam uji coba ini diberlakukan pembagian dataset sebesar 80% dataset *training* dan 20% dataset *testing*.

2.4 DATA TRANSFORM

Proses transformasi data adalah tahapan untuk mengorganisir data yang tergabung dalam sejumlah kelompok data menjadi satu himpunan kelompok data (Bahtiar, 2023). Langkah ini bertujuan untuk menyesuaikan bentuk dan cara penyajian data agar relevan ketika diterapkan pada analisis selanjutnya (Suriani, 2023). Dengan demikian, proses ini diharapkan dapat meningkatkan mutu data tersebut serta memaksimalkan wawasan yang dihasilkan saat diproses pada tahapan *data mining* mengenai jumlah kasus penderita penyakit untuk distribusi rumah sakit umum di tiap daerah di Provinsi Sumatera Selatan.

2.5 DATA MINING (PEMODELAN)

Setelah data melalui proses transformasi, penelitian ini masuk ke dalam tahapan pemodelan menggunakan algoritma K-Means untuk mengelompokkan wilayah berdasarkan variabel data yang telah dipilih. Algoritma K-Means tergolong dalam metode unsupervised learning yang sederhana dan lumrah digunakan dalam memecahkan solusi pengelompokan data. Algoritma ini termasuk dalam teknik non-hierarchy clustering yang memisah data menjadi kelompok-kelompok yang berbeda (Silvia Sasmita & Sri Winiarti, 2024). Keunggulan utama K-Means adalah kemudahan dalam implementasi, kesederhanaan, fleksibilitas, kecepatan, serta kerap diimplementasikan dalam proses penambangan data.

Menurut Hutagalung (2021), berikut tahapan dalam pengujian K-Means Clustering :

- Menentukan jumlah cluster (k) yang hendak dibangun.
- Menetapkan centroid awal dengan acak sebanyak k cluster.
- Mengkalculasi kedekatan masing-masing data ke tiap centroid dengan mengaplikasikan metrik Euclidean Distance hingga ditemukan centroid terdekat untuk setiap data.

Rumus Euclidean Distance:

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (1)$$

atau dalam bentuk penjumlahan eksplisit:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (2)$$

Keterangan:

$d(x,y)$ = Jarak antara objek data dan titik centroid.

x_k = Data ke- i pada atribut ke- k .

y_k = Nilai centroid pada atribut ke- k .

n = Jumlah atribut data.

- d) Menghitung ulang nilai centroid dengan mengambil rerata keseluruhan data dalam cluster tertentu.
- e) Mengulang tahapan perhitungan kedekatan pada masing-masing objek terhadap centroid yang telah diperbarui, selanjutnya tercantum di tahap (c).
- f) Mengulangi poin (c) dan (d) hingga tidak terjadi perubahan pada anggota cluster atau titik centroid mencapai kondisi stabil.

2.6 DATA EVALUATION

Algoritma K-Means digunakan dalam analisis data karena mampu mengelompokkan data berdasarkan kesamaan karakteristik dengan efisien serta memiliki kecepatan tinggi dalam memproses data berjumlah besar. Hasil interpretasi dari penggunaan algoritma K-Means dalam penelitian ini dapat menghasilkan pengetahuan mendalam serta pandangan baru mengenai pola distribusi jumlah kasus penderita berbagai jenis penyakit di berbagai wilayah di Sumatera Selatan. Dengan adanya pengelompokan ini, pemerintah dan instansi terkait dapat memahami wilayah mana yang memiliki tingkat kejadian penyakit lebih tinggi serta faktor-faktor yang berkontribusi terhadap persebarannya. Informasi ini sangat berguna dalam menentukan strategi yang lebih efektif untuk distribusi layanan rumah sakit umum, termasuk alokasi tenaga medis, fasilitas kesehatan, serta sumber daya lainnya. Selain itu, dengan pendekatan berbasis *knowledge discovery*, hasil clustering ini dapat dijadikan dasar dalam pengambilan keputusan berbasis data yang dapat dituangkan melalui kebijakan-kebijakan baru guna meningkatkan efisiensi dan efektivitas layanan kesehatan di wilayah yang membutuhkan perhatian lebih. Melalui penelitian ini, saran-saran inovasi yang dapat dikembangkan juga akan ditambahkan dengan pertimbangan logis berbasis hasil pemodelan data.

3. HASIL DAN PEMBAHASAN

Fase ini menjabarkan pandangan serta hasil dari tahapan penelitian sebelumnya menggunakan *tools* RapidMiner dan *Python Google Collaboratory* untuk mendapatkan kesimpulan terkait saran atau kebijakan dalam distribusi rumah sakit umum di setiap Kabupaten/Kota di Provinsi Sumatera Selatan.

3.1 DATA SELECTION

3.1.1 PEMILIHAN DATASET

Dataset terdiri dari 17 baris yang merepresentasikan kabupaten/kota di Provinsi Sumatera Selatan dan 21 kolom yang mencakup data tahunan (2019–2023) untuk empat atribut utama, yaitu jumlah penduduk, jumlah rumah sakit umum, jumlah penderita penyakit, dan jumlah tenaga kesehatan. Dataset ini terdiri dari 340 *record* yang berisi nilai dari atribut. Adapun *preview* data dari proses dataset selection terlihat pada Tabel 2.

Tabel 2. Dataset Selection

	A	B	C	D	E	F	G	H	I	J
	Kabupaten/Kota	Jumlah Penduduk 2019	Jumlah Penduduk 2020	Jumlah Penduduk 2021	Jumlah Penduduk 2022	Jumlah Penduduk 2023	Jumlah RSU 2019	Jumlah RSU 2020	Jumlah RSU 2021	Jumlah RSU 2022
1	Ogan Komering Ulu	368.756	367.603	371.106	375.538	379.130	4	4	4	4
2	Ogan Komering Ilir	832.151	769.348	772.742	797.429	797.429	3	3	3	3
3	Muara Enim	637.556	612.900	617.846	624.019	633.779	5	5	5	6
4	Lahat	409.348	430.071	434.939	441.174	444.949	2	2	2	2
5	Musi Rawas	405.175	395.570	398.732	402.674	407.694	2	2	2	2
6	Musi Banyuasin	649.085	622.206	627.070	633.124	644.386	3	3	3	3
7	Banyu Asin	857.097	836.914	843.871	852.576	874.210	4	5	5	6
8	Ogan Komering Ulu Selatan	363.004	408.981	416.616	426.687	424.190	1	1	1	1
9	Ogan Komering Ulu Timur	677.080	649.853	653.062	656.857	668.035	4	4	4	4
10	Ogan Ilir	429.595	416.549	419.401	422.907	431.043	3	3	3	3
11	Empat Lawang	250.465	333.622	343.839	357.673	342.178	2	2	2	2
12	Pali	190.062	194.900	197.290	200.368	202.681	1	1	1	1

3.1.2 PEMILIHAN FITUR ATAU VARIABEL

Variabel atau fitur yang dipilih untuk sampel data meliputi kabupaten/kota, jumlah penduduk, jumlah rumah sakit, jumlah penderita penyakit, dan jumlah tenaga kesehatan. Pada Tabel 3 ditunjukkan fitur-fitur yang akan dipilih untuk memperlihatkan korelasi antara penderita penyakit dengan ketersediaan fasilitas rumah sakit umum di kabupaten/kota provinsi Sumatera Selatan.

Tabel 3. Fitur Selection

	Kabupaten/ Kota (polynomial id)	Jumlah Penduduk 2019 (integer)	Jumlah Penduduk 2020 (integer)	Jumlah Penduduk 2021 (integer)	Jumlah Penduduk 2022 (integer)	Jumlah Penduduk 2023 (integer)
1	Ogan Komering Ulu	368.756	367.603	371.106	375.538	379.130
2	Ogan Komering Ilir	832.151	769.348	772.742	797.429	797.429
3	Muara Enim	637.556	612.900	617.846	624.019	633.779
4	Lahat	409.348	430.071	434.939	441.174	444.949
5	Musi Rawas	405.175	395.570	398.732	402.674	407.694
6	Musi Banyuasin	649.085	622.206	627.070	633.124	644.386
7	Banyu Asin	857.097	836.914	843.871	852.576	874.210
8	Ogan Komering Ulu Selatan	363.004	408.981	416.616	426.687	424.190
9	Ogan Komering Ulu Timur	677.080	649.853	653.062	656.857	668.035
10	Ogan Ilir	429.595	416.549	419.401	422.907	431.043
11	Empat Lawang	250.465	333.622	343.839	357.673	342.178
12	Pali	190.062	194.900	197.290	200.368	202.681

3.2 DATA PRE-PROCESSING

Pada tahap ini, dilakukan pembersihan dan persiapan data sebelum diterapkan algoritma K-Means. Langkah yang diimplementasikan pada tahap pre-processing adalah sebagai berikut.

3.2.1 PEMILIHAN DATA RELEVAN

Hanya data jumlah penderita penyakit (2019-2023) yang digunakan untuk analisis *clustering*, dengan Kabupaten/Kota tetap dipertahankan sebagai identifikasi daerah namun tidak dimasukkan dalam perhitungan kluster (data id) untuk keperluan data *testing*. Data relevan yang dipilih telah dijabarkan dan dapat dilihat pada Tabel 4.

Tabel 4. Pemilihan Data *Testing*

No.	Kabupaten/ Kota	Jumlah Penduduk 2019	Jumlah Penduduk 2020	Jumlah Penduduk 2021	Jumlah Penduduk 2022	Jumlah Penduduk 2023	Jumlah RSU 2019	Jumlah RSU 2020
1	Ogan Komering Ulu	368.756	367.603	371.106	375.538	379.130	4	4
2	Ogan Komering Ilir	832.151	769.348	772.742	797.429	797.429	3	3
3	Muara Enim	637.556	612.900	617.846	624.019	633.779	5	5
4	Lahat	409.348	430.071	434.939	441.174	444.949	2	2
5	Musi Rawas	405.175	395.570	398.732	402.674	407.694	2	2
6	Musi Banyuasin	649.085	622.206	627.070	633.124	644.386	3	3
7	Banyu Asin	857.097	836.914	843.871	852.576	874.210	4	5
8	Ogan Komering Ulu Selatan	363.004	408.981	416.616	426.687	424.190	1	1
9	Ogan Komering Ulu Timur	677.080	649.853	653.062	656.857	668.035	4	4
10	Ogan Ilir	429.595	416.549	419.401	422.907	431.043	3	3
11	Empat Lawang	250.465	333.622	343.839	357.673	342.178	2	2
12	Pali	190.062	194.900	197.290	200.368	202.681	1	1

3.2.2 PENANGANAN MISSING VALUES

Data diperiksa untuk memastikan tidak ada missing values yang dapat mempengaruhi hasil *clustering*, dan jika ditemukan, akan dilakukan imputasi atau penghapusan data sesuai kebutuhan. Dilakukan pembersihan data dengan menggunakan Python. Proses penanganan nilai yang hilang dimodelkan pada Gambar 2.

```
df_cleaned = df.drop(columns=["Unnamed: 21", "Unnamed: 22", "Unnamed: 23", "Unnamed: 24"])
```

Gambar 2. Penanganan *Missing Values*

3.2.3 FORMAT DATA KONSISTEN

Dipastikan semua data numerik memiliki format yang sesuai untuk diolah dalam model *machine learning*, dengan kolom tahun (2019-2023) dikonversi ke format numerik tanpa karakter tambahan. Lalu, bentuk kolom dari atribut-atribut yang akan digunakan harus diubah menjadi bentuk yang jelas agar dapat terbaca oleh sistem RapidMiner. Perubahan tersebut dapat diamati dalam Tabel 5 dan Tabel 6 di bawah ini.

Tabel 5. Kondisi Tabel Sebelum Dilakukan *Formatting*

Kabupaten/ Kota	Jumlah Penduduk					Kabupaten /Kota	Jumlah Rumah Sakit Umum				
	2019	2020	2021	2022	2023		2019	2020	2021	2022	2023
Ogan Komering Ulu	368.756	367.603	371.106	375.538	379.130	Ogan Komering Ulu	4	4	4	4	4
Ogan Komering Ilir	832.151	769.348	772.742	797.429	797.429	Ogan Komering Ilir	3	3	3	3	3
Muara Enim	637.556	612.900	617.846	624.019	633.779	Muara Enim	5	5	5	6	6

Lahat	409.3 48	430.0 71	434.9 39	441.1 74	444.9 49	Lahat	2	2	2	2	2
Musi Rawas	405.1 75	395.5 70	398.7 32	402.6 74	407.6 94	Musi Rawas	2	2	2	2	2
Musi Banyuasin	649.0 85	622.2 06	627.0 70	633.1 24	644.3 86	Musi Banyuasin	3	3	3	3	3
Banyu Asin	857.0 97	836.9 14	843.8 71	852.5 76	874.2 10	Banyu Asin	4	5	5	6	6
Ogan Komering Ulu Selatan	363.0 04	408.9 81	416.6 16	426.6 87	424.1 90	Ogan Komering Ulu Selatan	1	1	1	1	1
Ogan Komering Ulu Timur	677.0 80	649.8 53	653.0 62	656.8 57	668.0 35	Ogan Komering Ulu Timur	4	4	4	4	4
Ogan Ilir	429.5 95	416.5 49	419.4 01	422.9 07	431.0 43	Ogan Ilir	3	3	3	3	3
Empat Lawang	250.4 65	333.6 22	343.8 39	357.6 73	342.1 78	Empat Lawang	2	2	2	2	2
Pali	190.0 62	194.9 00	197.2 90	200.3 68	202.6 81	Pali	1	1	1	1	1
Musi Rawas Utara	192.5 40	188.8 61	190.4 20	192.4 69	195.9 62	Musi Rawas Utara	1	1	1	1	1
Kota Palembang	1.674. 243	1.668. 848	1.686. 073	1.707. 996	1.706. 371	Kota Palembang	23	23	23	22	23
Kota Prabumulih	188.6 69	193.1 96	195.7 48	199.0 47	200.6 73	Kota Prabumulih	4	4	4	4	4
Kota Pagaralam	139.1 92	143.8 44	145.2 66	147.0 71	147.8 36	Kota Pagaralam	1	1	1	1	1
Kota Lubuk Linggau	233.1 78	234.1 66	236.8 28	240.2 38	242.9 76	Kota Lubuk Linggau	3	4	4	4	4

Tabel 6. Kondisi Tabel Setelah Dirapikan

Kabupaten/Kota	Jumlah Penduduk tahun 2019	Jumlah Penduduk tahun 2020	Jumlah Penduduk tahun 2021	Jumlah Penduduk tahun 2022	Jumlah Penduduk tahun 2023
Ogan Komering Ulu	368756	367603	371106	375538	379130
Ogan Komering Ilir	832151	769348	772742	797429	797429
Muara Enim	637556	612900	617846	624019	633779
Lahat	409348	430071	434939	441174	444949
Musi Rawas	405175	395570	398732	402674	407694
Musi Banyuasin	649085	622206	627070	633124	644386
Banyu Asin	857097	836914	843871	852576	874210
Ogan Komering Ulu Selatan	363004	408981	416616	426687	424190
Ogan Komering Ulu Timur	677080	649853	653062	656857	668035
Ogan Ilir	429595	416549	419401	422907	431043
Empat Lawang	250465	333622	343839	357673	342178
Pali	190062	194900	197290	200368	202681
Musi Rawas Utara	192540	188861	190420	192469	195962
Kota Palembang	1674243	1668848	1686073	1707996	1706371
Kota Prabumulih	188669	193196	195748	199047	200673
Kota Pagaralam	139192	143844	145266	147071	147836
Kota Lubuk Linggau	233178	234166	236828	240238	242976

3.3 DATA TRANSFORM

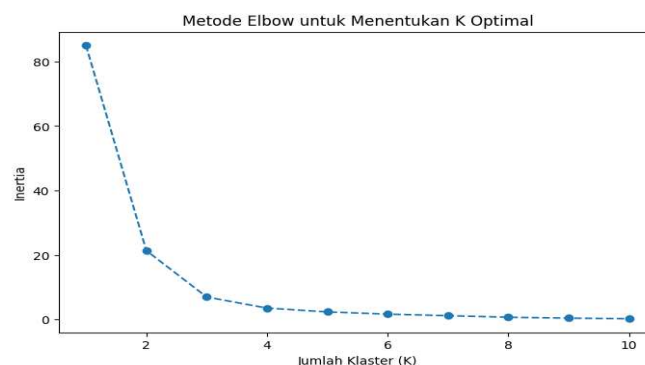
Pada tahap ini dilakukan normalisasi data sebelum dipindahkan ke *RapidMiner* dengan metode *StandardScaler*. Normalisasi yang dilakukan berupa mengubah skala fitur agar memiliki distribusi dengan rata-rata 0 dan standar revisi 1. Data yang dinormalisasi berupa data jumlah penderita penyakit tahun 2019, 2020, 2021, 2022, dan 2023. Pada kolom tahun dilakukan normalisasi sehingga kolom tahun ini berupa menjadi angka-angka terdistribusi. Hasil setelah normalisasi membuat kolom (tahun) memiliki skala yang seragam. Hal ini penting untuk memastikan bahwa setiap tahun (kolom) memiliki kontribusi yang setara dalam analisis kluster. Pada Gambar 3, digambarkan model dalam bahasa Python yang digunakan untuk menormalisasi data pada penelitian ini.

```
scaler = StandardScaler()
X_scaled = scaler.fit_transform(df_selected.iloc[:, 1:])
```

Gambar 3. Normalisasi *Standar Scaler*

3.4 DATA MINING

Di bagian ini diawali dengan metode Elbow dalam menentukan berapa total cluster yang optimal. Grafik Elbow yang ditunjukkan pada Gambar 4 menunjukkan titik penurunan inertia yang mulai melandai. Dari grafik tersebut, maka akan digunakan $K = 3$ untuk klasterisasi.



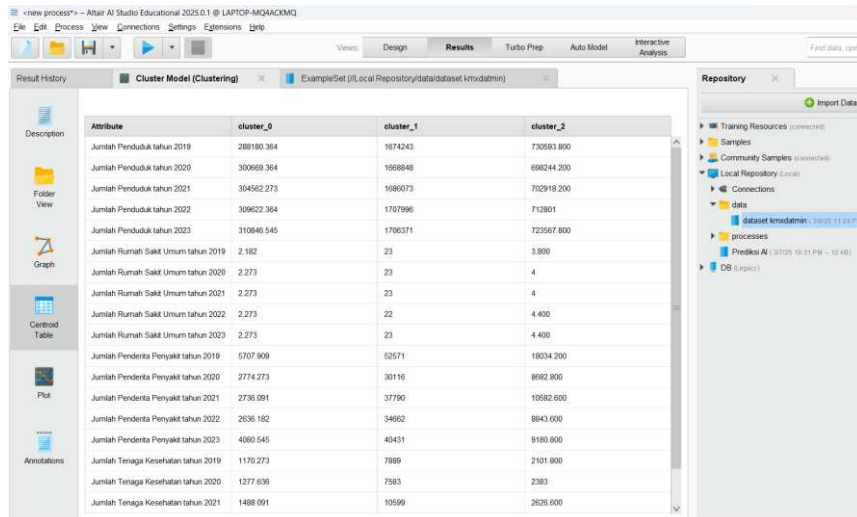
Gambar 4. Metode *Elbow*

Tiga *cluster* berbeda dihasilkan dari klasterisasi algoritma K-Means ini. Karakteristik dari setiap *cluster* berdasarkan variabel - variabel yang digunakan, seperti kabupaten/kota, jumlah penduduk, jumlah rumah sakit umum, jumlah kasus penderita penyakit, dan jumlah tenaga kesehatan. Dari proses klasterisasi menghasilkan tiga kelompok sebagai berikut. *Cluster 0* mempunyai 11 item, *Cluster 1* mempunyai 1 item, dan *Cluster 2* mempunyai 5 item. Model *Cluster* yang dihasilkan oleh algoritma K-Means ditunjukkan pada Tabel 7.

Tabel 7. Hasil Model Cluster Kabupaten/Kota Sumatera Selatan

Cluster 0 (Rendah)	Cluster 1 (Tinggi)	Cluster 2 (Sedang)
Ogan Komering Ulu	Kota Palembang	Ogan Komering Ilir
Lahat		Muara Enim
Musi Rawas		Musi Banyuasin
Ogan Komering Ulu Selatan		Banyuasin
Ogan Ilir		Ogan Komering Ulu Timur
Empat Lawang		
Pali		
Musi Rawas Utara		
Kota Prabumulih		
Kota Pagaralam		

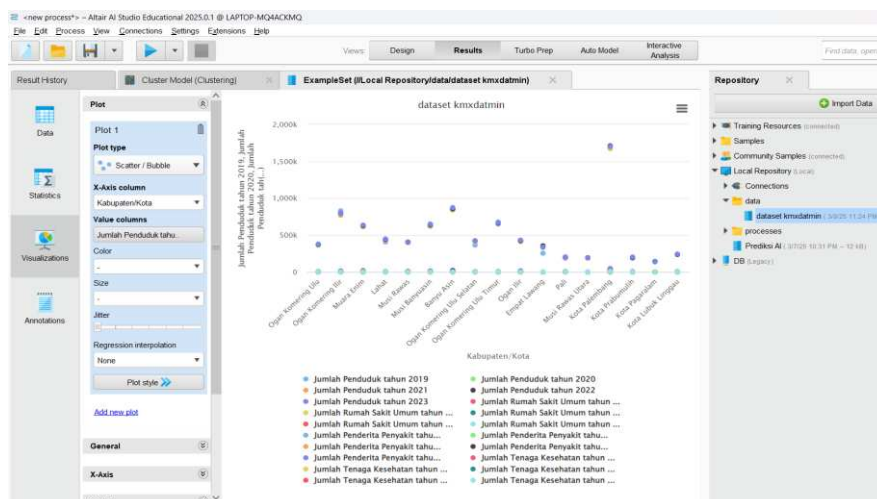
Setiap wilayah administratif (kabupaten/kota) telah diklasifikasikan masing-masing ke dalam satu cluster dari ketiga kluster yang ada. Pembagian kabupaten/kota terhadap kluster meliputi Cluster 0 (rendah) memiliki akses faskes yang rendah, Cluster 1 (tinggi) wilayah yang memiliki akses faskes yang memadai dan Cluster 2 (Sedang) wilayah yang memiliki akses faskes cukup memadai, yang hasilnya dapat diamati pada tabel diatas.



Attribute	cluster_0	cluster_1	cluster_2
Jumlah Penduduk tahun 2019	288190.364	1674243	730593.800
Jumlah Penduduk tahun 2020	300689.364	1668846	696244.200
Jumlah Penduduk tahun 2021	304562.273	1686073	702918.200
Jumlah Penduduk tahun 2022	306622.364	1707986	712801
Jumlah Penduduk tahun 2023	310846.545	1780371	723567.800
Jumlah Rumah Sakit Umum tahun 2019	2.182	23	3.800
Jumlah Rumah Sakit Umum tahun 2020	2.273	23	4
Jumlah Rumah Sakit Umum tahun 2021	2.273	23	4
Jumlah Rumah Sakit Umum tahun 2022	2.273	22	4.400
Jumlah Rumah Sakit Umum tahun 2023	2.273	23	4.400
Jumlah Penderita Penyakit tahun 2019	5707.909	62571	18034.200
Jumlah Penderita Penyakit tahun 2020	2774.273	30116	8992.800
Jumlah Penderita Penyakit tahun 2021	2736.091	37790	10562.600
Jumlah Penderita Penyakit tahun 2022	2636.182	34662	8943.600
Jumlah Penderita Penyakit tahun 2023	4080.545	40431	9180.800
Jumlah Tenaga Kesehatan tahun 2019	1170.273	7889	2101.800
Jumlah Tenaga Kesehatan tahun 2020	1277.636	7583	2363
Jumlah Tenaga Kesehatan tahun 2021	1488.091	10599	2626.800

Gambar 5. Tabel Centroid

Berdasarkan Gambar 5 dapat disimpulkan bahwa centroid data yang digunakan adalah sebagai berikut. *Cluster 1* (resiko tinggi) memiliki jumlah tenaga kesehatan paling banyak secara konsisten. Hal tersebut memungkinkan peningkatan tenaga kesehatan di wilayah ini untuk mengimbangi tingginya jumlah kasus penderita penyakit. Pada *cluster 2* (resiko sedang) juga mengalami peningkatan jumlah tenaga kesehatan, tetapi jumlahnya masih jauh lebih rendah dibandingkan *cluster 1*. Mengingat jumlah rumah sakit di *cluster 2* yang hanya sekitar 4 unit, memungkinkan hal ini mencerminkan keterbatasan infrastruktur kesehatan. *Cluster 0* (resiko rendah) memiliki tenaga kesehatan paling rendah. Peningkatannya sedikit dan konsisten, hal ini mencerminkan wilayah tersebut mungkin memiliki kebutuhan layanan kesehatan yang lebih rendah.



Gambar 6. Scatter Plot

Visualisasi hasil clustering menampilkan gambaran yang konkret mengenai penyebaran data. Scatter plot pada Gambar 6 menunjukkan hasil dari klasterisasi penyebaran data yang multidimensi. Dengan adanya visualisasi *Scatter plot* ini dapat mempermudah analisis dalam penyebaran data dan membantu mengidentifikasi model data. Berdasarkan statistik dari hasil clustering, penyebaran data dalam setiap *cluster* dapat disimpulkan sebagai berikut. *Cluster 0* mengandung kabupaten/kota dengan jumlah kasus penderita penyakit rendah. *Cluster 2* mengandung kabupaten/kota dengan jumlah kasus penderita penyakit sedang. Sementara *cluster 1* mengandung kabupaten/kota dengan jumlah kasus penderita penyakit sangat tinggi.

3.5 DATA EVALUATION (KNOWLEDGE DISCOVERY)

Berdasarkan tahap *data mining* yang sudah dilakukan sebelumnya, didapatkan hasil tiga buah klaster yang merepresentasikan wilayah kabupaten dan kota di Sumatera Selatan dengan tingkatan rendah, sedang, serta tinggi dari atribut yang digunakan sebagaimana dijelaskan pada Tabel 7 di bagian sebelumnya. Hasil dari klasterisasi tersebut mengindikasikan terdapat ketimpangan yang nyata antar wilayah, khususnya menyoroti Kota Palembang yang berada pada klaster 1 dengan wilayah lain pada klaster yang berbeda. Palembang memiliki jumlah rumah sakit dan tenaga kesehatan yang jauh lebih tinggi dibandingkan kabupaten lain, walaupun jumlah penderita penyakit di kota ini juga sangat besar. Sementara itu, *cluster* 0 dan 2 yang terdiri dari daerah dengan sumber daya kesehatan yang minim memerlukan perhatian lebih yang setara dalam pelayanan kesehatan.

Ketimpangan distribusi rumah sakit terlihat jelas dari hasil *clustering*, di mana hampir semua rumah sakit di Sumatera Selatan berpusat di Palembang, sementara wilayah lain terutama di *cluster* 0, memiliki keterbatasan akses layanan rumah sakit dengan rerata kurang lebih 2 rumah sakit per wilayah. Hal ini menunjukkan bahwa terdapat urgensi untuk membangun rumah sakit tipe C dan D yang berfokus pada layanan poliklinik terkhususnya pada penyakit IMS, HIV/AIDS, DBD, Diare, Malaria, dan Tuberkulosis di kabupaten dengan akses rumah sakit yang terbatas. Pemerintah daerah dapat membangun rumah sakit umum yang diselenggarakan dengan kebutuhan setiap daerah sehingga kebutuhan masyarakat dapat terpenuhi. Pembangunan fasilitas kesehatan ini juga melahirkan kesempatan untuk menjalin kolaborasi dengan CSR dari pihak swasta yang memiliki pengaruh di daerah-daerah tersebut. Dengan adanya sinergi antara pemerintah dan sektor swasta, diharapkan akses kesehatan dapat menjangkau masyarakat yang membutuhkan. Selain itu, keberadaan rumah sakit di *cluster* 2 dapat diperkuat dengan meningkatkan kapasitas layanan dengan mengoptimalkan jam pelayanan serta ruangan rawat inap agar pasien dari daerah sekitar tidak perlu dirujuk ke Palembang untuk mendapatkan akses kesehatan yang memadai.

Tabel 8. Nilai Centroid Clustering Rumah Sakit Umum dan Tenaga Kesehatan

Cluster	Centroid Rumah Sakit Umum					Centroid Tenaga Kesehatan				
	2019	2020	2021	2022	2023	2019	2020	2021	2022	2023
0	2.1	2.1	2.1	2.1	2.1	1190	1277.9	1487.1	1625.9	1649
1	23	23	23	22	23	7889	7583	10599	10546	11157
2	3.8	4	4	4.4	4.4	2101.8	2383	2626.6	2771.6	3083

Menilik dari hasil angka *centroid* dari proses *clustering* sebelumnya pada Tabel 8, distribusi rumah sakit bukanlah satu-satunya kebijakan yang mesti dibenahi. Angka penyebaran tenaga kesehatan pada wilayah di luar Kota Palembang juga memiliki selisih rerata yang cukup signifikan. Berdasarkan hasil pengujian *clustering* pada RapidMiner, nilai *centroid* untuk data tenaga kesehatan di tahun 2023 menampilkan bahwa *cluster* 0 memiliki rerata 6,7 kali lipat lebih rendah yakni 1649 orang dibandingkan *cluster* 1. Sedangkan pada *cluster* 2, rerata jumlah tenaga kesehatannya bernilai 3,6 kali lipat lebih rendah dengan jumlah 3083 orang dibandingkan *cluster* 1. Meski tingkat kepadatan penderita penyakit berbeda di setiap *cluster*, fakta ini menyoroti bahwa terdapat urgensi untuk melakukan pemerataan penyebaran tenaga kesehatan di wilayah-wilayah tersebut. Hal ini dinilai penting sebab layanan rumah sakit yang tersedia juga mesti dibersamai oleh tenaga kesehatan yang cukup dan kompeten. Pemerintah daerah dapat mengembangkan rencana untuk melakukan rotasi tenaga medis antar daerah dalam rangka memastikan bahwa dokter yang memiliki spesialisasi dalam bidangnya tidak hanya terpusat di Palembang, melainkan tersebar secara berkeadilan ke setiap wilayah lainnya. Tentunya, kebijakan ini juga perlu memperhatikan kesejahteraan tenaga medis yang ditugaskan ke daerah-daerah terpencil tersebut.

Merujuk kembali pada hasil *clustering* yang menyatakan bahwa terdapat daerah yang memiliki penyebaran rumah sakit umum dan tenaga kesehatan yang belum merata, dapat direkomendasikan pula optimalisasi layanan kesehatan berbasis teknologi. Inovasi ini dapat berupa *Mobile Health Clinic*, yakni klinik kesehatan bergerak dengan konsep kendaraan medis yang dilengkapi oleh alat-alat canggih untuk melakukan pemeriksaan dan pertolongan pertama bagi pasien darurat yang kesulitan untuk mengunjungi rumah sakit. Konsep *mobile health clinic* ini dapat diintegrasikan dengan *telemedicine station* berupa layanan kesehatan digital yang terhubung langsung dengan dokter yang berkaitan untuk dapat memberikan penanganan dan

diagnosis awal kepada pasien. Hal ini dapat dikembangkan dan diteliti lebih lanjut untuk mengoptimalkan resiko-resiko terkecil yang harus dihadapi oleh daerah pelosok akibat minimnya akses ke fasilitas kesehatan.

Urgensi hadirnya perbaikan kebijakan di atas selaras dengan adanya pernyataan di dalam Undang-Undang No.17 Tahun 2023 mengenai kesehatan, lalu pada pasal 7 ayat 1 bahwa Pemerintah Pusat dan Pemerintah Daerah bertanggung jawab untuk meningkatkan dan mengembangkan upaya kesehatan dalam rangka meningkatkan akses dan kualitas pelayanan kesehatan. Masyarakat juga memiliki hak mengacu pada pasal 4 ayat 1 poin c dan d yakni setiap orang berhak mendapatkan layanan kesehatan yang aman, bermutu, dan terjangkau agar dapat menciptakan derajat kesehatan yang setinggi-tingginya serta mendapat perawatan kesehatan sesuai dengan standar pelayanan kesehatan (Kementrian Kesehatan Republik Indonesia, 2023). Oleh karena itu, pemerintah perlu menindaklanjuti rekomendasi kebijakan ini dengan langkah konkret dan terencana agar kesenjangan layanan kesehatan di Sumatera Selatan dapat segera teratasi. Tanpa adanya komitmen dan implementasi yang serius, masyarakat di daerah dengan akses terbatas akan terus mengalami kesulitan dalam memperoleh layanan kesehatan yang maksimal.

4. SIMPULAN

Hasil clustering menunjukkan ketimpangan distribusi fasilitas kesehatan di Sumatera selatan. Kota Palembang (Cluster 1) memiliki jumlah rumah sakit dan tenaga medis jauh lebih tinggi dibanding daerah lain, dengan tenaga kesehatan 6,7 kali lebih banyak dibandingkan rata rata wilayah lain. Sebaliknya, wilayah dalam cluster 0 (rendah) hanya memiliki rata rata dua rumah sakit dan tenaga medis terbatas. Cluster (sedang) mengalami keterbatasan layanan seperti antrian panjang dan kurangnya fasilitas penunjang, yang masih menyebabkan ketergantungan pada Palembang. Ketimpangan ini berpotensi memperbesar ketidakadilan akses layanan kesehatan. Oleh karena itu, dibutuhkan pembangunan rumah sakit tipe C dan D di wilayah Cluster 0, peningkatan kapasitas rumah sakit di Cluster 2, serta pemertaan tenaga medis melalui rotasi dokter spesialis. Inovasi seperti Mobile Health Clinic dan telemedicine juga penting untuk menjangkau daerah terpencil. Langkah ini mendukung amanat UU No. 17 Tahun 2023 tentang kesehatan. Temuan ini dapat menjadi dasar bagi pengambil kebijakan dalam merumuskan intervensi yang lebih merata dan berbasis data.

5. DAFTAR PUSTAKA

- Atalya, M., Leza, A., Utami, W., Anugrah, P., & Dewi, C. (2024). Prediksi Prestasi Siswa Smas Katolik Santo Yoseph Denpasar Berdasarkan Kedisiplinan Dan Tingkat Ekonomi Orang Tua Menggunakan Metode Knowledge Discovery in Database dan Algoritma Regresi Linier Berganda. *Jurnal Mahasiswa Teknik Informatika*, 8(1). <https://doi.org/https://doi.org/10.36040/jati.v8i1.8754>
- Badan Pusat Statistik Provinsi Sumatera Selatan. (2023). *Jumlah kasus penderita penyakit (kasus)*. <https://sumsel.bps.go.id/id/statistics-table/2/Mzc1IzI=/jumlah-kasus-penderita-penyakit-kasus-.html>
- Badan Pusat Statistik Provinsi Sumatera Selatan. (2023). *Jumlah rumah sakit umum, rumah sakit khusus, puskesmas, klinik pratama, dan posyandu menurut kabupaten/kota di Provinsi Sumatera Selatan*. <https://sumsel.bps.go.id/id/statistics-table/3/YmlzemNGUkNVblZLVVhObIREWnZXbkEzWld0eVVUMDkjMw==/jumlah-rumah-sakit-umum--rumah-sakit-khusus--puskesmas--klinik-pratama--dan-posyandu-menurut-kabupaten-kota-di-provinsi-sumatera-selatan.html?year=2023>
- Badan Pusat Statistik Provinsi Sumatera Selatan. (2023). *Jumlah tenaga kesehatan menurut kabupaten/kota di Provinsi Sumatera Selatan*. <https://sumsel.bps.go.id/id/statistics-table/3/YVdwSFJHRjRVVkJqWIRWRU9EQkhNVFY0UjB4VVVUMDkjMw==/jumlah-tenaga-kesehatan-menurut-kabupaten-kota-di-provinsi-sumatera-selatan.html?year=2023>
- Bahtiar, R. (n.d.). Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor. *Jurnal Informatika MULTI*, 1(3), 203–214. <https://jurnal.publikasitecno.id/index.php/jim203>
- Bahtiar, R. (2023). Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor. *Jurnal Informatika MULTI*, 1(3), 203–214. <https://jurnal.publikasitecno.id/index.php/multi/article/view/33>
- Dinas Kesehatan Provinsi Sumatera Selatan. (2023). (2023). *Profil Kesehatan Provinsi Sumatera Selatan Tahun 2023*. <https://dinkes.sumselprov.go.id/>
- Gesicho, M. B., Were, M. C., & Babic, A. (2021). Evaluating performance of health care facilities at meeting HIV-indicator reporting requirements in Kenya: an application of K-means clustering algorithm. *BMC Medical Informatics and Decision Making*, 21(1). <https://doi.org/10.1186/s12911-020-01367-9>

- Hendriansyah, B. A., Tri, A., Harjanta, J., & Latifah, K. (2025). Implementasi Algoritma K-Means Clustering Pada Sistem Informasi Geografis Fasilitas Kesehatan Bpjs Kesehatan Kota Semarang. *Jurnal Informatika Teknologi Dan Sains (JINTEKS)*, 27(1), 438–448. <https://doi.org/https://doi.org/10.51401/jinteks.v7i1.5618>
- Hutagalung, J., & Sonata, F. (2021). Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(3), 1187. <https://doi.org/10.30865/mib.v5i3.3113>
- Kementerian Kesehatan Republik Indonesia. (2023). *Undang-Undang Nomor 17 Tahun 2023 tentang Kesehatan*. <https://farmalkes.kemkes.go.id/unduh/uu-17-2023/>
- Kristina Sitinjak, D., Aji Pangestu, B., & Nurina Sari, B. (2022). Clustering Tenaga Kesehatan Berdasarkan Kecamatan di Kabupaten Karawang Menggunakan Algoritma K-Means. *Journal Of Applied Informatics and Computing*, 27(1), 438–448. <https://doi.org/https://doi.org/10.30871/jaic.v6i1.3855>
- Mayola, L., Hafizhb, M., & Syahputra, H. S. (2025). Klasterisasi Rumah Sakit berdasarkan Kunjungan Pasien menggunakan Algoritma K-Means: Data 2019-2023. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 7(1), 15–21. <https://doi.org/10.47233/jteksis.v7i1.1703>
- Minarni, M., Sari, E. I., Syahrani, A., & Mandarani, P. (2021). Klasterisasi Penyakit Menggunakan Algoritma K-Medoids pada Dinas Kesehatan Kabupaten Agam. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 10(3), 137. <https://doi.org/10.23887/janapati.v10i3.34904>
- Mohammed, M., & Alsunosi, R. (2022). Effect of Selecting Validation Dataset on Building Random Forest and Decision Tree Models. *Med App Sci*, 5(2), 470–478. <https://doi.org/10.5281/zenodo.7113928>
- Nirwana Sevi Dian. (2022). *Perbandingan Algoritma K-Means Dan K-Medoids Dalam*. Universitas Sriwijaya.
- Ordila, R., Wahyuni, R., Irawan, Y., & Yulia Sari, M. (2020). Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus : Poli Klinik PT.Inecda). *Jurnal Ilmu Komputer*, 9(2), 148–153. <https://doi.org/10.33060/jik/2020/vol9.iss2.181>
- Permatasari, N. E., & Kristiyanti, L. (2023). Faktor-faktor yang Memengaruhi Kesejahteraan Masyarakat Di Desa Buntar. *Jurnal Ilmiah Keuangan Akuntansi Bisnis*, 2(3), 420–429. <https://doi.org/10.53088/jikab.v2i3.68>
- Sekar Setyaningtyas, Indarmawan Nugroho, B., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-MeanS. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 10(2), 52–61. <https://doi.org/10.21063/jtif.2022.v10.2.52-61>
- Silvia Sasmita, M., & Sri Winiarti, I. (2024). Implementasi Algoritma K-Means Untuk Clustering Data Penyakit Di Puskesmas Kotagede 2 Yogyakarta. *SUBMIT (Jurnal Ilmiah Teknologi Informasi Dan Sains*, 4(1), 1–9. <https://doi.org/https://doi.org/10.36815/submit.v4i1.3195>
- Suriani, U. (2023). Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5. *Journal of Computer and Information Systems Ampera*, 3(2). <https://doi.org/10.51519/journalcisa.v4i2.393>
- Utomo, Y. B., Kurniasari, I., & Yanuartanti, I. (2023). Penerapan Knowledge Discovery In Database Untuk Analisa Tingkat Kecelakaan Lalu Lintas. *Jurnal Teknik Informatika Kaputama (JTik)*, 7(1). <https://doi.org/https://doi.org/10.59697/jtik.v7i1.61>
- V. Novalia, K. Ditha Tania, A. Meiriza, & A. Wedhasmara. (2024). Knowledge Discovery of Application Review Using Word Embedding's Comparison with CNN-LSTM Model on Sentiment Analysis. *2024 International Conference on Electrical Engineering and Computer Science (ICECOS)*, 234–238. [https://doi.org/EngineeringandComputerScience\(ICECOS\),Palembang,Indonesia,2024,pp.234-238,doi:10.1109/ICECOS63900.2024.10791113](https://doi.org/EngineeringandComputerScience(ICECOS),Palembang,Indonesia,2024,pp.234-238,doi:10.1109/ICECOS63900.2024.10791113)
- Vira, S. (2021). Pengelompokan Provinsi di Indonesia Berdasarkan Penyakit Tidak Menular Menggunakan Metode Partisi, Hierarki, dan Fuzzy Clustering Salsavira. *Jurnal Matematika Dan Statistika Serta Aplikasinya*, 9(1), 94–103. <https://doi.org/https://doi.org/10.24252/msa.v9i1.17942>
- Yulianti, Y., & Qomariah, S. (2025). Indeks Pembangunan Manusia Ilmu Pengetahuan. *CENDEKIA: Jurnal Ilmu Sosial, Bahasa Dan Pendidikan*, 5(1), 203–213. <https://doi.org/10.55606/cendekia.v5i1.3512>